Country Music Project

Matt Shu

2022-05-03

## Prereqs

# Needed to overcome error found below with Homebrew TBB vs bundled TBB:   
# https://github.com/RcppCore/RcppParallel/issues/182  
# remotes::install\_github("RcppCore/RcppParallel")   
library(igraph)  
library(tidyverse)  
library(stm)  
library(RSQLite)  
library(RecordLinkage)  
library(stringdist)  
library(devtools)  
library(tm)  
devtools::install\_github("mikajoh/tidystm", dependencies = TRUE)  
library(tidystm)

conn <- dbConnect(RSQLite::SQLite(), "files/22-04-21-playback-fm-top-country.db")  
cleaned\_df <- dbGetQuery(conn, 'SELECT \* FROM cleaned')  
dbDisconnect(conn)

names(cleaned\_df)

## [1] "artist\_id" "track\_id" "year"   
## [4] "artist" "track" "rank"   
## [7] "link" "lyrics" "artist\_appearances"  
## [10] "mb\_id" "type" "area.name"   
## [13] "gender" "life\_span.begin" "life\_span.ended"   
## [16] "song\_id" "cleaned\_lyrics" "lyrics\_alnum"

cleaned\_df <- cleaned\_df %>%   
 mutate(gender = replace(gender, gender == "other", "non-binary")) %>%   
 mutate(gender = replace(gender, is.na(gender), "group"))

## Preprocessing (and STM exploration)

# Dataframe containing the text  
docs\_df <- cleaned\_df %>%  
 dplyr::select(track\_id, lyrics\_alnum) %>%  
 # first, remove observation with missing values of the meta variables  
 filter(!is.na(lyrics\_alnum)) %>%  
 # the objects need to be class "data frame"   
 as.data.frame()

# Dataframe containing (sample) documents' metadata of interest  
meta\_df <- cleaned\_df %>%  
 dplyr::select(track\_id, rank, artist, track, year, gender) %>%  
 # the objects need to be class "data frame"   
 as.data.frame()

processed\_docs\_1 <- textProcessor(documents = docs\_df$lyrics\_alnum,   
 metadata = meta\_df,   
 lowercase = TRUE,   
 removestopwords = TRUE,   
 removenumbers = TRUE,   
 removepunctuation = TRUE,   
 ucp = TRUE,  
 stem = TRUE,   
 striphtml = TRUE,   
 wordLengths = c(3, Inf),  
 language = "en")

meta <- processed\_docs\_1$meta  
vocab <- processed\_docs\_1$vocab  
docs <- processed\_docs\_1$documents  
keep <- !is.na(meta$artist) && !is.na(meta$rank)

## Warning in !is.na(meta$artist) && !is.na(meta$rank): 'length(x) = 5970 > 1' in  
## coercion to 'logical(1)'  
  
## Warning in !is.na(meta$artist) && !is.na(meta$rank): 'length(x) = 5970 > 1' in  
## coercion to 'logical(1)'

meta <- meta[keep,]  
docs <- docs[keep]

prepped\_data <- prepDocuments(docs,   
 vocab,   
 meta,  
 # the lower threshold value means that only words  
 # that appear more times than the value (in this   
 # example the value = 3) will be retained; this is   
 # another researcher decision  
 lower.thresh = 2)

Old code for removing unusual mismatch with no words despite past filters

length(docs\_df$lyrics\_alnum) # original documents  
length(prepped\_data$meta$track\_id) # off from the preceding count  
dif <- setdiff(docs\_df$track\_id, # original vector of documents  
 prepped\_data$meta$track\_id) # list of documents after prepDocuments  
tmp <- docs\_df  
tmp2 <- tmp[!tmp$track\_id %in% dif,]  
tmp\_doc <- tmp2 %>%  
 select(track\_id, lyrics\_alnum)  
length(tmp\_doc$track\_id)  
length(prepped\_data$meta$track\_id)  
  
# View the track ids that were removed for some reason (often other language)  
tmp3 <- tmp[tmp$track\_id %in% dif,]  
tmp3

See Cleaned Sample!

head(cleaned\_df)

## artist\_id track\_id year artist track rank  
## 1 1 0 1944 Red Foley Smoke On The Water 1  
## 2 1 506 1951 Red Foley Hobo Boogie 55  
## 3 1 587 1953 Red Foley Midnight 14  
## 4 1 386 1950 Red Foley Cincinnati Dancing Pig 13  
## 5 1 374 1950 Red Foley Chattanoogie Shoe Shine Boy 1  
## 6 1 620 1953 Red Foley Hot Toddy 47  
## link  
## 1 /charts/country/video/1944/red-foley-smoke-on-the-water  
## 2 /charts/country/video/1951/red-foley-hobo-boogie  
## 3 /charts/country/video/1953/red-foley-midnight  
## 4 /charts/country/video/1950/red-foley-cincinnati-dancing-pig  
## 5 /charts/country/video/1950/red-foley-chattanoogie-shoe-shine-boy  
## 6 /charts/country/video/1953/red-foley-hot-toddy  
## lyrics  
## 1 Smoke On The Water LyricsThere will be a sad day comin' For the foes of all mankind They must answer to the people And it’s troubling their mind Everybody who must fear them Will rejoice on that great day When the powers of dictators Shall be taken all away There’ll be smoke on the water On the land and the sea When our Army and Navy overtakes the enemy There’ll be smoke on the mountains Where the Heathen Gods stay And the sun that is risin’ Will go down on that day For there is a great destroyer Made of fire and flesh and steel Rollin’ toward the foes of freedom They’ll go down beneath its wheels There’ll be nothing left but vultures To inhabit all that land When our modern ships and bombers Make a graveyard of Japan Hirohito ‘long with Hitler Will be ridin’ on a rail Mussolini’ll beg for mercy As a leader he has failed But there’ll be no time for pity When the Screamin’ Eagle flies That will be the end of Axis They must answer with their lives Embed  
## 2 Hobo Boogie Lyrics For a sandwich or a quarter, or a spider something wet We will play that old piano like you have never heard it yet The hobo boogie, that what he calls his melody Got the rhythm in a boxcar ridin' on a bumpy rail Wrote the tune from start to finish on the walls of the county jail The hobo boogie, that's just the way it seem to be He tires of faces, everyday He tires of places, wants to get away He gets the urge to hit the steel To Asbill, Nashville, Chattanooga, or Mobil You can hear the wheels a-clangin' and the rails just seem to hum It'll make you feel like shoutin' "Hey there Birmingham, here I come!" The hobo boogie, you got to love that melody He tires of faces, everyday He tires of places, wants to get away He gets the urge to hit the steel To Asbill, Nashville, Chattanooga, or Mobil You can hear the wheels a-clangin' and the rails just seem to hum It'll make you feel like shoutin' "Hey there Birmingham, here I come!" The hobo boogie, you got to love that melodyEmbed  
## 3 Midnight LyricsMidnight, I've spent another lonely day, thinkin' of you Midnight, tomorrow is on it's way, empty and blue I'm so lonely, so lonely at midnight for you Midnight, oh what a lonely time to weep, I ought to know Midnight, I should have been fast asleep hours ago Still I'm crying, I'm crying 'cause I miss you so Midnight, I lie in bed awake and stare at nothing at all Wonderin', wonderin' why you don't care, wishing you'd call Tears keep flowing like drops from a waterfall Midnight, I've spent another lonely day, thinkin' of you Midnight, tomorrow is on it's way, empty and blue I'm so lonely, so lonely at midnight for youEmbed  
## 4 Cincinnati Dancing Pig Lyrics Cincinnati's dancin' pig, he's the barnyard's mister big Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Children yell and clap and sing When he does his buckin' wing Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Dancin' bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain't-a seen nothin' at all From Duluth to Birmingham, soo wee He's the porkchop Dapper Dan He's the keenest hog white am, Cincinnati's dancin' pig With his riggidy-jiggidy-jiggidy-jig-jig-jig Cincinnati's dancin' pig, he's the barnyard's mister big Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Children yell and clap and sing When he does his buckin' wing Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Dancin' bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain't-a seen nothin' at all From Duluth to Birmingham, soo wee He's the porkchop Dapper Dan He's the keenest hog white am, Cincinnati's dancin' pig With his riggidy, with his jiggidy, jig-jig-jigEmbed  
## 5 Chattanoogie Shoe Shine Boy LyricsHave you ever passed the corner of Forth and Grand? Where a little ball o' rhythm has a shoe-shine stand People gather 'round and they clap their hands He's a great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy He charges you a nickel just to shine one shoe He makes the oldest kind o' leather look like new You feel as though you wanna dance when he gets through He's a great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy It's a wonder that the rag don't tear The way he makes it pop You ought to see him fan the air With his hoppity-hippity-hippity-hoppity-hoppity-hippity-hop He opens up for business when the clock strikes nine He likes to get up early when they're feelin' fine Everybody gets a little rise 'n shine With the great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy It's a wonder that the rag don't tear The way he makes it pop Just listen to him fan the air Here he goes! He opens up for business when the clock strikes nine He likes to get up early when they're feelin' fine Everybody gets a little rise 'n shine With the great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy The Chattanoogie shoe-shine boyEmbed  
## 6 Hot Toddy LyricsHot Toddy sure makes a body feel mighty nice, mighty nice One swallow of a hot Hot Toddy is worth any price, any price Two swallows of a hot Hot toddy will break the ice, break the ice Three swallows of a hot Hot toddy will add the spice Four swallows of a hot Hot toddy, you're thinkin' twice, thinkin' twice Five swallows of a hot Hot toddy, they're throwin' rice, throwin' rice You're needin' some hot Hot toddy, take my advice, my advice Start cookin' with hot Hot toddy and coast to paradise Hot Toddy sure makes a body feel mighty nice, mighty nice One swallow of a hot Hot Toddy is worth any price, any price Two swallows of a hot Hot toddy will break the ice, break the ice Three swallows of a hot Hot toddy will add the spice Four swallows of a hot Hot toddy, you're thinkin' twice, thinkin' twice Five swallows of a hot Hot toddy, they're throwin' rice, throwin' rice You're needin' some hot Hot toddy, take my advice, my advice Start cookin' with hot Hot toddy and coast to paradise Coast to paradise Coast to paradiseEmbed  
## artist\_appearances mb\_id type area.name  
## 1 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 2 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 3 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 4 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 5 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 6 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## gender life\_span.begin life\_span.ended song\_id  
## 1 male 1910-06-17 true 14519  
## 2 male 1910-06-17 true 11892  
## 3 male 1910-06-17 true 13445  
## 4 male 1910-06-17 true 10833  
## 5 male 1910-06-17 true 10810  
## 6 male 1910-06-17 true 11966  
## cleaned\_lyrics  
## 1 There will be a sad day comin' For the foes of all mankind They must answer to the people And it’s troubling their mind Everybody who must fear them Will rejoice on that great day When the powers of dictators Shall be taken all away There’ll be smoke on the water On the land and the sea When our Army and Navy overtakes the enemy There’ll be smoke on the mountains Where the Heathen Gods stay And the sun that is risin’ Will go down on that day For there is a great destroyer Made of fire and flesh and steel Rollin’ toward the foes of freedom They’ll go down beneath its wheels There’ll be nothing left but vultures To inhabit all that land When our modern ships and bombers Make a graveyard of Japan Hirohito ‘long with Hitler Will be ridin’ on a rail Mussolini’ll beg for mercy As a leader he has failed But there’ll be no time for pity When the Screamin’ Eagle flies That will be the end of Axis They must answer with their lives Embed  
## 2 For a sandwich or a quarter, or a spider something wet We will play that old piano like you have never heard it yet The hobo boogie, that what he calls his melody Got the rhythm in a boxcar ridin' on a bumpy rail Wrote the tune from start to finish on the walls of the county jail The hobo boogie, that's just the way it seem to be He tires of faces, everyday He tires of places, wants to get away He gets the urge to hit the steel To Asbill, Nashville, Chattanooga, or Mobil You can hear the wheels a-clangin' and the rails just seem to hum It'll make you feel like shoutin' "Hey there Birmingham, here I come!" The hobo boogie, you got to love that melody He tires of faces, everyday He tires of places, wants to get away He gets the urge to hit the steel To Asbill, Nashville, Chattanooga, or Mobil You can hear the wheels a-clangin' and the rails just seem to hum It'll make you feel like shoutin' "Hey there Birmingham, here I come!" The hobo boogie, you got to love that melodyEmbed  
## 3 Midnight, I've spent another lonely day, thinkin' of you Midnight, tomorrow is on it's way, empty and blue I'm so lonely, so lonely at midnight for you Midnight, oh what a lonely time to weep, I ought to know Midnight, I should have been fast asleep hours ago Still I'm crying, I'm crying 'cause I miss you so Midnight, I lie in bed awake and stare at nothing at all Wonderin', wonderin' why you don't care, wishing you'd call Tears keep flowing like drops from a waterfall Midnight, I've spent another lonely day, thinkin' of you Midnight, tomorrow is on it's way, empty and blue I'm so lonely, so lonely at midnight for youEmbed  
## 4 Cincinnati's dancin' pig, he's the barnyard's mister big Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Children yell and clap and sing When he does his buckin' wing Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Dancin' bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain't-a seen nothin' at all From Duluth to Birmingham, soo wee He's the porkchop Dapper Dan He's the keenest hog white am, Cincinnati's dancin' pig With his riggidy-jiggidy-jiggidy-jig-jig-jig Cincinnati's dancin' pig, he's the barnyard's mister big Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Children yell and clap and sing When he does his buckin' wing Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Dancin' bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain't-a seen nothin' at all From Duluth to Birmingham, soo wee He's the porkchop Dapper Dan He's the keenest hog white am, Cincinnati's dancin' pig With his riggidy, with his jiggidy, jig-jig-jigEmbed  
## 5 Have you ever passed the corner of Forth and Grand? Where a little ball o' rhythm has a shoe-shine stand People gather 'round and they clap their hands He's a great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy He charges you a nickel just to shine one shoe He makes the oldest kind o' leather look like new You feel as though you wanna dance when he gets through He's a great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy It's a wonder that the rag don't tear The way he makes it pop You ought to see him fan the air With his hoppity-hippity-hippity-hoppity-hoppity-hippity-hop He opens up for business when the clock strikes nine He likes to get up early when they're feelin' fine Everybody gets a little rise 'n shine With the great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy It's a wonder that the rag don't tear The way he makes it pop Just listen to him fan the air Here he goes! He opens up for business when the clock strikes nine He likes to get up early when they're feelin' fine Everybody gets a little rise 'n shine With the great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy The Chattanoogie shoe-shine boyEmbed  
## 6 Hot Toddy sure makes a body feel mighty nice, mighty nice One swallow of a hot Hot Toddy is worth any price, any price Two swallows of a hot Hot toddy will break the ice, break the ice Three swallows of a hot Hot toddy will add the spice Four swallows of a hot Hot toddy, you're thinkin' twice, thinkin' twice Five swallows of a hot Hot toddy, they're throwin' rice, throwin' rice You're needin' some hot Hot toddy, take my advice, my advice Start cookin' with hot Hot toddy and coast to paradise Hot Toddy sure makes a body feel mighty nice, mighty nice One swallow of a hot Hot Toddy is worth any price, any price Two swallows of a hot Hot toddy will break the ice, break the ice Three swallows of a hot Hot toddy will add the spice Four swallows of a hot Hot toddy, you're thinkin' twice, thinkin' twice Five swallows of a hot Hot toddy, they're throwin' rice, throwin' rice You're needin' some hot Hot toddy, take my advice, my advice Start cookin' with hot Hot toddy and coast to paradise Coast to paradise Coast to paradiseEmbed  
## lyrics\_alnum  
## 1 There will be a sad day comin For the foes of all mankind They must answer to the people And it s troubling their mind Everybody who must fear them Will rejoice on that great day When the powers of dictators Shall be taken all away There ll be smoke on the water On the land and the sea When our Army and Navy overtakes the enemy There ll be smoke on the mountains Where the Heathen Gods stay And the sun that is risin Will go down on that day For there is a great destroyer Made of fire and flesh and steel Rollin toward the foes of freedom They ll go down beneath its wheels There ll be nothing left but vultures To inhabit all that land When our modern ships and bombers Make a graveyard of Japan Hirohito long with Hitler Will be ridin on a rail Mussolini ll beg for mercy As a leader he has failed But there ll be no time for pity When the Screamin Eagle flies That will be the end of Axis They must answer with their lives Embed  
## 2 For a sandwich or a quarter or a spider something wet We will play that old piano like you have never heard it yet The hobo boogie that what he calls his melody Got the rhythm in a boxcar ridin on a bumpy rail Wrote the tune from start to finish on the walls of the county jail The hobo boogie that s just the way it seem to be He tires of faces everyday He tires of places wants to get away He gets the urge to hit the steel To Asbill Nashville Chattanooga or Mobil You can hear the wheels a clangin and the rails just seem to hum It ll make you feel like shoutin Hey there Birmingham here I come The hobo boogie you got to love that melody He tires of faces everyday He tires of places wants to get away He gets the urge to hit the steel To Asbill Nashville Chattanooga or Mobil You can hear the wheels a clangin and the rails just seem to hum It ll make you feel like shoutin Hey there Birmingham here I come The hobo boogie you got to love that melodyEmbed  
## 3 Midnight I ve spent another lonely day thinkin of you Midnight tomorrow is on it s way empty and blue I m so lonely so lonely at midnight for you Midnight oh what a lonely time to weep I ought to know Midnight I should have been fast asleep hours ago Still I m crying I m crying cause I miss you so Midnight I lie in bed awake and stare at nothing at all Wonderin wonderin why you don t care wishing you d call Tears keep flowing like drops from a waterfall Midnight I ve spent another lonely day thinkin of you Midnight tomorrow is on it s way empty and blue I m so lonely so lonely at midnight for youEmbed  
## 4 Cincinnati s dancin pig he s the barnyard s mister big Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Children yell and clap and sing When he does his buckin wing Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Dancin bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain t a seen nothin at all From Duluth to Birmingham soo wee He s the porkchop Dapper Dan He s the keenest hog white am Cincinnati s dancin pig With his riggidy jiggidy jiggidy jig jig jig Cincinnati s dancin pig he s the barnyard s mister big Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Children yell and clap and sing When he does his buckin wing Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Dancin bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain t a seen nothin at all From Duluth to Birmingham soo wee He s the porkchop Dapper Dan He s the keenest hog white am Cincinnati s dancin pig With his riggidy with his jiggidy jig jig jigEmbed  
## 5 Have you ever passed the corner of Forth and Grand Where a little ball o rhythm has a shoe shine stand People gather round and they clap their hands He s a great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy He charges you a nickel just to shine one shoe He makes the oldest kind o leather look like new You feel as though you wanna dance when he gets through He s a great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy It s a wonder that the rag don t tear The way he makes it pop You ought to see him fan the air With his hoppity hippity hippity hoppity hoppity hippity hop He opens up for business when the clock strikes nine He likes to get up early when they re feelin fine Everybody gets a little rise n shine With the great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy It s a wonder that the rag don t tear The way he makes it pop Just listen to him fan the air Here he goes He opens up for business when the clock strikes nine He likes to get up early when they re feelin fine Everybody gets a little rise n shine With the great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy The Chattanoogie shoe shine boyEmbed  
## 6 Hot Toddy sure makes a body feel mighty nice mighty nice One swallow of a hot Hot Toddy is worth any price any price Two swallows of a hot Hot toddy will break the ice break the ice Three swallows of a hot Hot toddy will add the spice Four swallows of a hot Hot toddy you re thinkin twice thinkin twice Five swallows of a hot Hot toddy they re throwin rice throwin rice You re needin some hot Hot toddy take my advice my advice Start cookin with hot Hot toddy and coast to paradise Hot Toddy sure makes a body feel mighty nice mighty nice One swallow of a hot Hot Toddy is worth any price any price Two swallows of a hot Hot toddy will break the ice break the ice Three swallows of a hot Hot toddy will add the spice Four swallows of a hot Hot toddy you re thinkin twice thinkin twice Five swallows of a hot Hot toddy they re throwin rice throwin rice You re needin some hot Hot toddy take my advice my advice Start cookin with hot Hot toddy and coast to paradise Coast to paradise Coast to paradiseEmbed

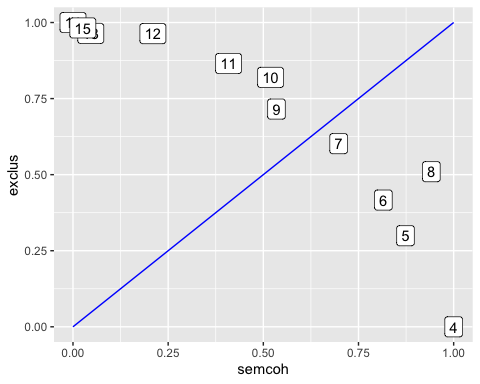
### Find K

k\_seq = seq(4, 15, 1)

## You can "watch" the algorithm model topics in the console  
searched = searchK(prepped\_data$documents,  
 prepped\_data$vocab,  
 K = k\_seq,  
 data = prepped\_data$meta,   
 seed = 183654)  
saveRDS(searched, file = "files/22-04-29-searchK.RData")

### Show K

searched <- readRDS("files/22-04-29-searchK.RData")  
# Get values from `searchK` output  
semcoh <- unlist(searched$results$semcoh)  
exclus <- unlist(searched$results$exclus)  
  
# Max/min semantic cohesion  
max\_sc <- max(semcoh)  
min\_sc<-min(semcoh)  
  
# Max/min exclusivity  
max\_ex<-max(exclus)  
min\_ex<-min(exclus)  
  
# Min-max normalization is (value - min)/(max - min)  
x\_vals <- (semcoh-min\_sc)/(max\_sc-min\_sc)  
y\_vals <- (exclus-min\_ex)/(max\_ex-min\_ex)  
# add semantic cohesion and exclusivity together weighted evenly  
search\_plot\_df <- tibble(id = k\_seq,   
 semcoh = x\_vals,  
 exclus = y\_vals,   
 combine = x\_vals\*0.5 + y\_vals\*0.5)  
# Plot  
ggplot(search\_plot\_df, mapping = aes(x = semcoh, y = exclus)) +  
 xlim(0,1) +  
 ylim(0,1) +  
 ggplot2::annotate("segment", x = 0, xend = 1, y = 0, yend = 1, color = "blue") +  
 geom\_label(aes(label=id))



unique(prepped\_data$meta$gender)

## [1] "male" "group" "female" "non-binary"

### Model Work

# 6 topics seems to also work nice, with a strong "Country" category  
num\_topics <- 7 # Chosen after above search and some playing around  
out\_covariates\_7 <- stm(prepped\_data$documents,  
 prepped\_data$vocab,  
 K = num\_topics,  
 prevalence = ~ rank + year \* gender,  
 max.em.its = 500,  
 data = prepped\_data$meta,  
 seed = 592669)

head(out\_covariates\_7$theta) # each row is each document

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0.183253946 0.007069806 0.017951578 0.14437433 0.034003629 0.351454015  
## [2,] 0.059540942 0.026539450 0.028958227 0.64243815 0.174697316 0.040460910  
## [3,] 0.465211708 0.018478813 0.071580114 0.01325611 0.089937091 0.305363030  
## [4,] 0.006668523 0.007024786 0.008411345 0.90506787 0.004507195 0.031606214  
## [5,] 0.040222095 0.034340099 0.068512467 0.40076420 0.030625939 0.198152810  
## [6,] 0.032443538 0.613071173 0.034814173 0.01632440 0.038018646 0.009980522  
## [,7]  
## [1,] 0.26189269  
## [2,] 0.02736501  
## [3,] 0.03617314  
## [4,] 0.03671407  
## [5,] 0.22738239  
## [6,] 0.25534755

# To find each artists, link the songs to the artists and then take the average for each artists, for each artist average of the columns  
head(prepped\_data$meta) # same order between dataframes

## track\_id rank artist track year gender  
## 1 0 1 Red Foley Smoke On The Water 1944 male  
## 2 506 55 Red Foley Hobo Boogie 1951 male  
## 3 587 14 Red Foley Midnight 1953 male  
## 4 386 13 Red Foley Cincinnati Dancing Pig 1950 male  
## 5 374 1 Red Foley Chattanoogie Shoe Shine Boy 1950 male  
## 6 620 47 Red Foley Hot Toddy 1953 male

track\_topic\_df <- cbind(prepped\_data$meta, out\_covariates\_7$theta)

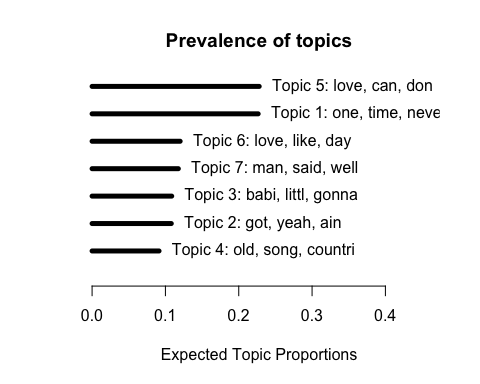
terms = labelTopics(out\_covariates\_7, n = 10)  
terms$prob # rows are topics; columns are most probable words (in order)

## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]   
## [1,] "one" "time" "never" "now" "heart" "still" "say" "just" "gone"  
## [2,] "got" "yeah" "ain" "like" "girl" "good" "get" "wanna" "just"  
## [3,] "babi" "littl" "gonna" "come" "time" "night" "get" "take" "back"  
## [4,] "old" "song" "countri" "roll" "back" "town" "road" "ride" "sing"  
## [5,] "love" "can" "don" "know" "just" "want" "let" "make" "feel"  
## [6,] "love" "like" "day" "dream" "night" "eye" "blue" "sweet" "rain"  
## [7,] "man" "said" "well" "old" "daddi" "boy" "big" "mama" "just"  
## [,10]   
## [1,] "cri"   
## [2,] "can"   
## [3,] "right"  
## [4,] "like"   
## [5,] "need"   
## [6,] "light"  
## [7,] "got"

terms$frex # rows are topics; columns are most FREX words (in order)

## [,1] [,2] [,3] [,4] [,5] [,6] [,7]   
## [1,] "fool" "goodby" "cri" "lone" "heartach" "memori" "tear"   
## [2,] "ooh" "huh" "boo" "yeah" "whoa" "nothin" "ain"   
## [3,] "bye" "babi" "honey" "bit" "gonna" "shake" "danc"   
## [4,] "countri" "boogi" "hillbilli" "cowboy" "crank" "cha" "tonk"   
## [5,] "want" "hold" "need" "lose" "fall" "love" "believ"   
## [6,] "angel" "rain" "heaven" "sail" "sea" "storm" "rainbow"  
## [7,] "mom" "dad" "wife" "hero" "father" "twenti" "america"  
## [,8] [,9] [,10]   
## [1,] "miss" "lie" "still"   
## [2,] "nobodi" "lovin" "woah"   
## [3,] "step" "littl" "batter"  
## [4,] "jone" "tennesse" "doo"   
## [5,] "easi" "give" "don"   
## [6,] "sunshin" "wing" "sky"   
## [7,] "daddi" "sir" "famili"

# Parameters modified from: https://milesdwilliams15.github.io/Better-Graphics-for-the-stm-Package-in-R/  
par(bty="n",lwd=5)  
plot(out\_covariates\_7,  
 type = "summary",  
 main = "Prevalence of topics")



docs\_examples\_covar <- findThoughts(out\_covariates\_7,  
 texts = tmp\_doc$track\_id,  
 n = 10,  
 topics = c(1:num\_topics))  
  
for(topic\_num in c(1:num\_topics)) {  
 print(paste("Topic ", topic\_num))  
 for(track in docs\_examples\_covar$docs[[topic\_num]]) {  
 print(cleaned\_df$track[cleaned\_df$track\_id == track])  
 }  
 print("")  
}

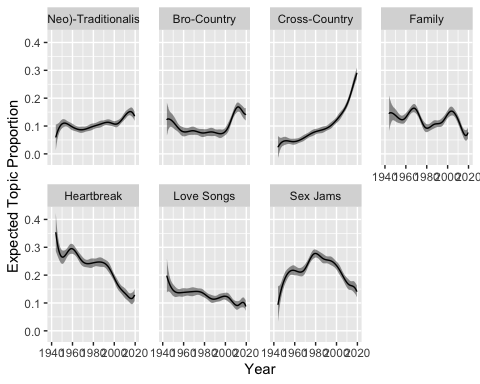
## [1] "Topic 1"  
## [1] "Rag Mop"  
## [1] "Something Old, Something New"  
## [1] "I Forgot To Remember To Forget"  
## [1] "All Alone in This World without You"  
## [1] "Fool Fool Fool"  
## [1] "Happy Journey"  
## [1] "Careless Darlin'"  
## [1] "Sweetheart You Done Me Wrong"  
## [1] "You're The One"  
## [1] "Things Aren't Funny Anymore"  
## [1] ""  
## [1] "Topic 2"  
## [1] "Desperate Man"  
## [1] "Gimmie That Girl"  
## [1] "My Bucket's Got a Hole in it"  
## [1] "Just the Way"  
## [1] "Just The Way"  
## [1] "Uh-Huh--Mm"  
## [1] "Uh-Huh-mm"  
## [1] "She Ain't Your Ordinary Girl"  
## [1] "Glad You Exist"  
## [1] "Drinkin' Beer. Talkin' God. Amen."  
## [1] ""  
## [1] "Topic 3"  
## [1] "Swing"  
## [1] "Trademark"  
## [1] "Little Bit of Life"  
## [1] "Little Bit Of Life"  
## [1] "Baby Let's Play House"  
## [1] "Penny Arcade"  
## [1] "Whole Lotta Shakin' Goin' On"  
## [1] "Shine, Shave, Shower (It's Saturday)"  
## [1] "Be-Bop-A-Lula"  
## [1] "Last Minute Late Night"  
## [1] ""  
## [1] "Topic 4"  
## [1] "Teenage Boogie"  
## [1] "Redneck Yacht Club"  
## [1] "Cincinnati Dancing Pig"  
## [1] "Ragtime Cowboy Joe"  
## [1] "Long Live"  
## [1] "Mule Train"  
## [1] "She Cranks My Tractor"  
## [1] "The Rhumba Boogie"  
## [1] "Smokey Mountain Boogie"  
## [1] "Hula Rock"  
## [1] ""  
## [1] "Topic 5"  
## [1] "Love Can't Wait"  
## [1] "Don't Underestimate My Love For You"  
## [1] "Don't Underestimate My Love for You"  
## [1] "I Want To Know You Before We Make Love"  
## [1] "Count on Me"  
## [1] "A Lover's Question"  
## [1] "Mr. Lovemaker"  
## [1] "It Matters to Me"  
## [1] "It Matters To Me"  
## [1] "Fall into Me"  
## [1] ""  
## [1] "Topic 6"  
## [1] "Ring Of Fire"  
## [1] "Your Name Is Beautiful"  
## [1] "Sweet Summer Lovin'"  
## [1] "Mockin' Bird Hill"  
## [1] "The Red Strokes"  
## [1] "A Fallen Star"  
## [1] "Would You Lay With Me (In A Field Of Stone)"  
## [1] "Kentucky Waltz"  
## [1] "Beautiful Brown Eyes"  
## [1] "My Special Angel"  
## [1] ""  
## [1] "Topic 7"  
## [1] "What's Your Mama's Name"  
## [1] "Life Of A Poor Boy"  
## [1] "(Margie's At) The Lincoln Park Inn"  
## [1] "No Charge"  
## [1] "History Repeats Itself"  
## [1] "Poor, Poor Pitiful Me"  
## [1] "Deck Of Cards"  
## [1] "Po' Folks"  
## [1] "Shiftwork"  
## [1] "None Of My Business"  
## [1] ""

# Topic 1: Heartbreak Songs  
# Topic 2: Cross-Country (Country Rock/Pop)  
# Topic 3: Traditionalist Country (Pardi, Hank Williams)  
# Topic 4: Bro-Country  
# Topic 5: Sex Jams  
# Topic 6: Love songs  
# Topic 7: Family  
topic\_labels <- c("Heartbreak", "Cross-Country", "(Neo)-Traditionalist", "Bro-Country", "Sex Jams", "Love Songs", "Family")

eff1 <- estimateEffect(formula = c(1:num\_topics) ~ s(year),  
 # the line above matches the model specification we used  
 stmobj = out\_covariates\_7,  
 meta = prepped\_data$meta,  
 uncertainty = "Global")  
  
# plot.estimateEffect(eff1,  
# covariate = "year",  
# topics = c(1:num\_topics),  
# model = out\_covariates\_7,  
# method = "continuous",  
# xlab = "Year",  
# ylim=c(0, .4),  
# xlim=c(1940, 2020),  
# main = "Effect of Year on Topic Proportion")

effect <- lapply(c(0, 1), function(i) {  
 extract.estimateEffect(eff1,  
 covariate = "year",  
 topics = c(1:num\_topics),  
 model = out\_covariates\_7,  
 method = "continuous")  
})  
effect <- do.call("rbind", effect)  
effect <- effect %>% mutate(label = recode(topic, "1"=topic\_labels[1], "2" = topic\_labels[2], "3" = topic\_labels[3], "4" = topic\_labels[4], "5" = topic\_labels[5], "6" = topic\_labels[6], "7" = topic\_labels[7]))  
## And, for example, plot it with ggplot2 and facet by topic instead.  
library(ggplot2)  
  
ggplot(effect, aes(x = covariate.value, y = estimate,  
 ymin = ci.lower, ymax = ci.upper)) +  
 facet\_wrap(~ label, nrow = 2) +  
 geom\_ribbon(alpha = .5) +  
 geom\_line() +  
 labs(x = "Year",  
 y = "Expected Topic Proportion") +   
 scale\_x\_continuous(breaks=c(1940, 1960, 1980, 2000, 2020),   
 labels=waiver(), lim=c(1940,2020)) +   
 theme(panel.spacing = unit(1, "lines"))

## Warning: Removed 4 row(s) containing missing values (geom\_path).



eff <- estimateEffect(formula = c(1:num\_topics) ~ year,  
 # the line above matches the model specification we used  
 stmobj = out\_covariates\_7,  
 meta = prepped\_data$meta,  
 uncertainty = "Global")  
  
# Second, plot the results  
plot(eff,  
 covariate = "year",  
 topics = c(1:num\_topics),  
 model = out\_covariates\_7,  
 method = "continuous",  
 xlab = "Year",  
 main = "Effect of Year on Topic Proportion")

library(huge)

## Registered S3 methods overwritten by 'huge':  
## method from  
## plot.sim lava  
## print.sim lava

topic\_corr <- topicCorr(out\_covariates\_7, method = "huge")

## Conducting the nonparanormal (npn) transformation via shrunkun ECDF....done.  
## Conducting Meinshausen & Buhlmann graph estimation (mb)....done  
## Conducting rotation information criterion (ric) selection....done  
## Computing the optimal graph....done

topic\_corr

## $posadj  
## 7 x 7 sparse Matrix of class "dgCMatrix"  
##   
## [1,] . . . . 1 . .  
## [2,] . . 1 1 . . .  
## [3,] . 1 . 1 . . .  
## [4,] . 1 1 . . . 1  
## [5,] 1 . . . . . .  
## [6,] . . . . . . .  
## [7,] . . . 1 . . .  
##   
## $poscor  
## 7 x 7 sparse Matrix of class "dgCMatrix"  
##   
## [1,] . . . . 0.03444665 . .   
## [2,] . . 0.138984635 0.052890434 . . .   
## [3,] . 0.13898463 . 0.002650995 . . .   
## [4,] . 0.05289043 0.002650995 . . . 0.07017762  
## [5,] 0.03444665 . . . . . .   
## [6,] . . . . . . .   
## [7,] . . . 0.070177618 . . .   
##   
## $cor  
## 7 x 7 Matrix of class "dgeMatrix"  
## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 0.00000000 -0.34030035 -0.282685092 -0.383999053 0.03444665 -0.1442173  
## [2,] -0.34030035 0.00000000 0.138984635 0.052890434 0.00000000 -0.2355104  
## [3,] -0.28268509 0.13898463 0.000000000 0.002650995 0.00000000 -0.2104287  
## [4,] -0.38399905 0.05289043 0.002650995 0.000000000 -0.41828640 -0.1191443  
## [5,] 0.03444665 0.00000000 0.000000000 -0.418286396 0.00000000 0.0000000  
## [6,] -0.14421725 -0.23551039 -0.210428750 -0.119144260 0.00000000 0.0000000  
## [7,] 0.00000000 0.00000000 -0.132888013 0.070177618 -0.35784909 -0.1596963  
## [,7]  
## [1,] 0.00000000  
## [2,] 0.00000000  
## [3,] -0.13288801  
## [4,] 0.07017762  
## [5,] -0.35784909  
## [6,] -0.15969634  
## [7,] 0.00000000  
##   
## attr(,"class")  
## [1] "topicCorr"

set.seed(5)  
png('figures/topic\_corr.png')  
plot(topic\_corr,   
 vlabels = topic\_labels, vertex.label.cex = 0.8, layout = layout.auto)  
dev.off()

## quartz\_off\_screen   
## 2

Topics 3, 2, 4, 7 are all related. This is an interesting finding! This suggests that traditionalist country especially seems related to both country rock/pop songs Topic 2?: Country Rock/Pop Topic 3: Traditionalist Country Topic 4: Bro-Country Topic 7: Family

## More on Topic Models

### Questions/Interests

* How would I see where individual artists fell in terms of topics?
* In general, seeing prevalence of certain
* Would it be, taking the top x documents for different topics and counting from there? ### More to Do?
* Plot covariate interaction!
  + Particularly interested in tracking gender \* year interactions!